Surgical Endoscopy (2019) 33:2162–2168
https://doi.org/10.1007/s00464-018-6495-3

Using virtual-reality simulation to ensure basic competence in hysteroscopy

Mona M. Savran[1] · Anders Bo Nielsen[2] · Bente Baekholm Poulsen[3] · Poul Bak Thorsen[4] · Lars Konge[2,5]

Received: 19 May 2018 / Accepted: 11 October 2018 / Published online: 17 October 2018

Abstract

Background Hysteroscopy is a technically challenging procedure. Specialty curricula of obstetrics and gynaecology appraise hysteroscopy for trainees but there is no present evidence-based training program that certifies the fundamental technical skills before performance on patients. The objectives of this study were to develop and gather validity evidence for a simulation-based test that can ensure basic competence in hysteroscopy.

Methods We used the virtual-reality simulator HystMentor™. Six experts evaluated the feasibility and clinical relevance of the simulator modules. Six modules were selected for the test and a pilot study was carried out. Subsequently, medical students, residents, and experienced gynaecologists were enrolled for testing. Outcomes were based on generated simulator metrics. Validity evidence was explored for all five sources of evidence (content, response process, internal structure, relations to other variables, consequences of testing).

Results Inter-case reliability was high for four out of five metrics (Cronbach's alpha ≥ 0.80). Significant differences were identified when comparing the three groups' performances (p values < 0.05). Participants' clinical experience was significantly correlated to their simulator test score (Pearson's $r = 0.49$, $p < 0.001$). A single medical student managed to achieve the established pass/fail score (6.7% false positive) and three experienced gynaecologists failed the test (27.3% false negative).

Conclusions We developed a virtual-reality simulation-based test in hysteroscopy with supporting validity evidence. The test is intended to ensure competency in a mastery learning program where trainees practise on the simulator until they are able to pass before they proceed to supervised training on patients.

Keywords Hysteroscopy · Gynaecology · Education · Virtual reality · Simulation training

Hysteroscopy is a common procedure with a broad range of both diagnostic and operative indications including visualisation of the uterine cavity and treatment of various pathologies such as menorrhagia, polyps, leiomyomas, and suspicion of endometrial abnormalities [1]. Biopsy specimens are also sampled with hysteroscopy in the diagnosis of endometrial cancer [2].

As a minimally invasive procedure, hysteroscopy is associated with fewer complications than traditional open surgery [3]. However, it is challenging to gain manual dexterity in hysteroscopy because of the fulcrum effect and the orientation between the two-dimensional screen and the three-dimensional uterine cavity. Other difficulties are the diminished tactile feedback and limited degrees of freedom [4]. Lack of proficiency in hysteroscopy can compromise patient safety by causing uterine perforation and trauma to neighbouring organs. Also, pulmonary air embolisms and pulmonary oedema are potential life-threatening complications

✉ Mona M. Savran
monasavran@gmail.com

1 Department of Obstetrics and Gynecology, Copenhagen University Hospital Amager and Hvidovre, Kettegård Alle 30, 2650 Hvidovre, Capital Region of Denmark, Denmark

2 SimC, Odense University Hospital, Odense, Region of Southern Denmark, Denmark

3 Department of Obstetrics and Gynecology, Odense University Hospital, Odense, Region of Southern Denmark, Denmark

4 Research Unit for Gynecology and Obstetrics, Department of Clinical Research, University of Southern Denmark, Odense, Denmark

5 Center for HR, Copenhagen Academy for Medical Education and Simulation (CAMES), Copenhagen, Capital Region of Denmark, Denmark

[5]. The complexity of performing hysteroscopy reckons the necessity of training [6]. A recently published systematic review on hysteroscopic training and assessment demonstrated scarce research with merely 26 eligible studies [7]. In that particular study sample, authors unanimously agreed that training in hysteroscopy lead to a rise of qualifications; however, the components of training varied immensely. More importantly, the studies were supported by limited validity evidence.

It is recommended that residents in obstetrics and gynaecology gain competence in hysteroscopy and several curricula have been suggested [8, 9]. Simulation-based training with virtual reality offers a great opportunity to achieve the technical skills of hysteroscopy in a safe and controlled learning environment prior to clinical practice on patients [10]. The training can be tailored by using the approach of mastery learning, which is key for sound competence. The prospect of mastery learning is to reach an intended level of proficiency without any constraints on training duration. Compared with non-mastery training, this approach is correlated to improved skills acquisition [11].

A test with a predetermined pass/fail score is indispensable for implementing a mastery learning training program [12]. Testing must consequently be based on ample evidence of validity in order to correctly give meaning to test results. Systematically gathered validity arguments are critical in supporting test interpretation, which is particularly imperative for high-stakes exams as seen in residency programs [13]. To our knowledge, such a test does not exist in the discipline of hysteroscopy.

The objectives of this study were to develop and gather validity evidence for a simulation-based test that can ensure basic competence in hysteroscopy before performance of live surgery.

Materials and methods

Simulator modules and simulator metrics

The study was conducted with the simulator HystMentor™ (3D Systems, South Carolina) containing pre-installed diagnostic and operative modules of hysteroscopy. The diagnostic modules test access gaining, application of distension media, visualisation of a retroverted uterus, navigation, and biopsy sampling. The operative modules are polypectomy and myomectomy. The complexity of the modules is differentiated by easy, medium, difficult, and advanced. Every module is recorded by several metrics including resected pathology, coagulated surface, obtained/lost biopsy samples, distention media, duration of obscured view, duration of collapsed uterus, active tool without contact, active tool when pushing, cervix contacts, cavity contacts, visualisation of the uterine cavity/left ostium/right ostium, procedure time, cumulative path length, and camera alignment. Every metric is measured by an integrated scoring system that generates a dimensionless quantity representing the individual score. In addition, the metrics are measured in SI units.

Test content

A collaboration of two simulator experts (MMS, ABN), one professor in medical education (LK), two experienced gynaecologists (BBP, PBT), and one experienced hysteroscopist (performed > 2000 hysteroscopies) evaluated the feasibility and clinical relevance of the diagnostic and operative simulator modules in a Delphi-like. All members of the panel were instructed to assess the metrics and the modules and were requested to prioritise those with high clinical applicability. The test should consist of several modules with increasing difficulty and be possible to complete in maximum 45 min by an experienced gynaecologist. This single-round agreement yielded a test of six relevant modules: polypectomy easy, polypectomy medium, myomectomy medium, myomectomy difficult, advanced resection multiple polyps, and advanced resection multiple myomas. Subsequently, the test was programmed in the simulator software. The members of the panel were excluded from test participation. A pilot test was performed with four participants of different experience levels counting two final-year medical students, one resident with limited experience in hysteroscopy (performed < 50 hysteroscopies), and one experienced gynaecologist (performed > 50 hysteroscopies). They all completed the six modules in a satisfactory manner and the final test was decided upon.

Participants

Three groups of distinct experience levels were enrolled. Medical students were defined as final-year students with no hands-on hysteroscopy experience and were recruited from the University of Southern Denmark; residents with limited hysteroscopy experience (1–50 hysteroscopy procedures); and experienced gynaecologists, who had performed more than 50 hysteroscopies. Participants were excluded if they had prior practice with virtual-reality hysteroscopy simulation. All physicians were from the Departments of Obstetrics and Gynaecology, University Hospitals of Southern Denmark. We intended to include more than 10 participants per group which allows a presumption of normally distributed data in medical education research [14].

Testing

The enrolment took place from April 2017 to November 2017. Each individual was granted a unique identification number for test participation to preserve their anonymity. They also filled out a paper on participant characteristics notifying the approximate number of performed hysteroscopies. An identical introductory session of 10 min was provided for the participants. In this session, they were thoroughly instructed on how to use the simulator instruments and they then practised the following diagnostic modules: access gaining (2 attempts); application of distension media (1 attempt); visualisation of a retroverted uterus (1 attempt); and navigation (1 attempt). Subsequently, the test of six modules was administered to the participants who were supervised during testing and the supervisor (ABN) used the same instructions in writing for each participant. Their test answers were only accessible by the main investigator (MMS) and the 2nd author (ABN).

Outcome measure

An algorithm was constructed based on the total scores on the simulator metrics chosen by the panel to generate an overall score for each participant:

$$\text{Overall score} = \frac{\text{Resected pathology (\%)} \times \text{Camera alignment (\%)}}{\text{Procedure time (minutes)} \times \text{distension fluid (litres)} \times \text{path length (metres)}}$$

A high overall score would be obtained if the camera was aligned and all pathology was removed in an efficient way, that is, without spending too much time, using too much distension fluid, or making unnecessary instrument movements thus increasing the path length.

The included metrics in SI units were consistently measured in all six modules. The software-generated dimensionless scores were omitted due to their unspecified derivation.

Data analysis

The contemporary framework of validity by Messick was used to establish validity evidence of the test [15, 16]. Five sources of validity outline this framework:

Content

The test content had been evaluated by a panel of experienced members as described previously.

Response process

Consistent instructions for instrument handling and identical introductory simulator modules were provided by the same instructor (ABN) prior to testing. Performances were measured exclusively by the simulator and all test scores were stored in an SPSS file and analysed consistently. The simulator software was not updated during data collection.

Internal structure

Cronbach's alpha was calculated to explore the internal consistency of the test. The measures were inter-case reliabilities using the recorded simulator metrics as cases.

Relations to other variables

Based on performances on each metric (mean values) and on the algorithm-derived overall scores, the three groups were compared with each other by conducting ANOVA tests with Bonferroni corrections (post hoc analysis). A Pearson's correlation coefficient was obtained to examine the correlation between Overall Scores and experience in hysteroscopy.

Consequences of testing

A pass/fail score for the test was identified by the Contrasting Groups' Method [17]; two distributions were plotted based on the mean overall scores and the standard deviations of the medical students and experienced gynaecologists. The intersection of the two distributions was defined as the pass/fail score.

Statistical analysis was carried out in SPSS Statistics version 24.0 (IBM, NY, USA). p values less than 0.05 were considered as statistically significant.

Results

A total of 43 participants counting 15 medical students, 17 residents, and 11 experienced gynaecologists completed the test.

Internal structure

Inter-case reliability was calculated using Cronbach's alpha for each of the five metrics in the six modules. High coefficients were found for camera alignment: 0.81; procedure time: 0.85; distention media: 0.81; and path length: 0.80. The coefficient for the metric resected pathology was 0.58.

Table 1 The three group test performances according to performed hysteroscopies

Groups	Number of performed hysteroscopies	Mean overall score	SD	95% CI for mean Lower bound	95% CI for mean Upper bound	Minimum score	Maximum score
Medical students (*n* = 15)	0	4184	4054	1939	6429	407	12,809
Residents (*n* = 17)	1–35	17,109	14,549	9629	24,590	1733	48,708
Experienced gynaecologists (*n* = 11)	50–2000	31,090	24,462	14 657	47,523	5329	87,316

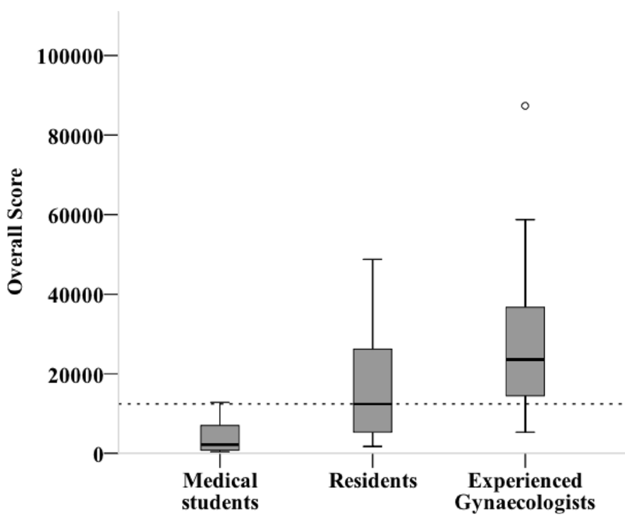

Fig. 1 Box-and-whiskers plot showing the overall scores for the three groups. The median, the minimum scores, and the maximum scores are also depicted. The dotted line demonstrates the pass/fail score

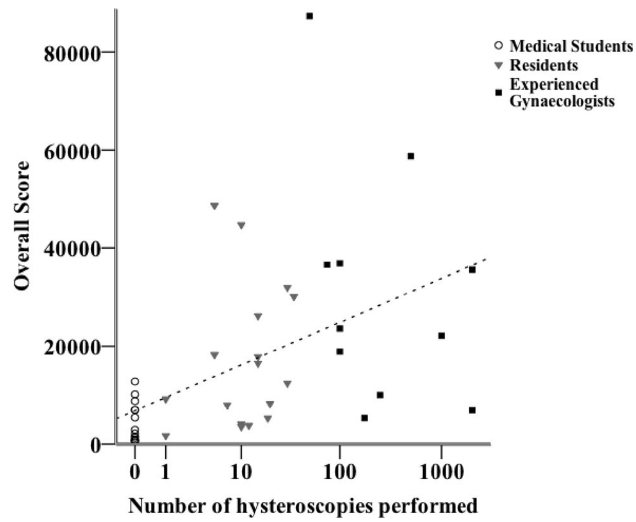

Fig. 2 Graph demonstrating a linear association between the logarithm of hysteroscopy experience and the overall score for the three groups

Relations to other variables

One-way ANOVA produced statistically significant differences for all of the five metrics (resected $\text{PATHOLOGY}_{\text{TOTAL}}$ $p = 0.004$; camera $\text{alignment}_{\text{Total}}$ $p = 0.04$; procedure $\text{time}_{\text{Total}}$ $p < 0.001$; distention $\text{fluid}_{\text{Total}}$ $p = 0.001$; path $\text{length}_{\text{Total}}$ $p < 0.001$), and for the overall scores $p < 0.001$ (Table 1). The test performances are demonstrated in Fig. 1. Post hoc analysis by applying the Bonferroni correction showed statistically significant differences in overall score between the medical students and the experienced gynaecologists ($p < 0.001$); however, no differences were found between the medical students and residents ($p = 0.07$) and between the residents and the experienced gynaecologists ($p = 0.07$). The logarithm of experience in hysteroscopy was statistically significant correlated with the Overall Scores; the Pearson correlation coefficient was 0.49 ($p < 0.001$) (Fig. 2).

Consequences of testing

The pass/fail score using the intersection between the performances of the medical students and experienced gynaecologists was accordingly set at the overall score 12,400 (Fig. 3). Thus, a single medical student passed (6.7% false positive), whilst three experienced gynaecologists failed the test (27.3% false negative). Nine of the residents (53%) scored above the pass/fail score.

Discussion

We developed a simulation-based test on hysteroscopy with supporting evidence from all five sources of the contemporary framework of validity.

Six operative simulator modules and five associated metrics were deemed relevant by an assorted panel of experts (validity evidence regarding *content*). Opinion sampling of experts has been shown beneficial in several other studies

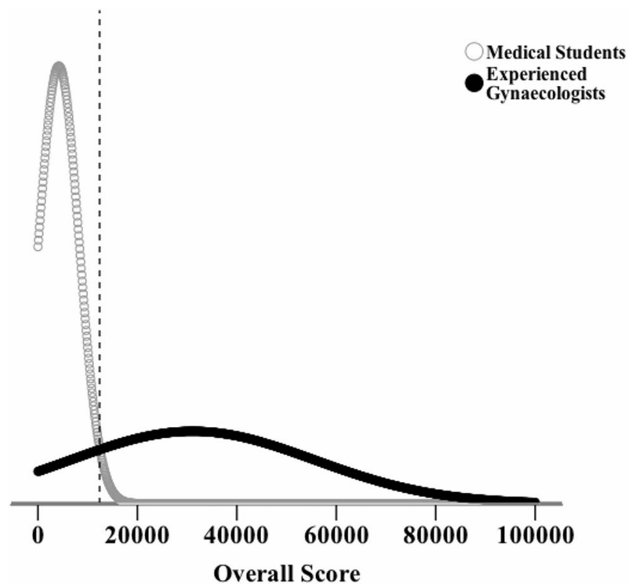

Fig. 3 Pass/fail score obtained by the Contrasting Groups' method is illustrated by the dashed line

to obtain consensus on test content or other areas that lack well-defined features [18]. This approach is inspired by the Delphi method in which expert ratings on a certain matter are summed up anonymously during a number of rating rounds till consensus is reached [19]. Similarly, the experts in our study did not discuss their opinions with each other, though only a single round was required to settle. One might address subjectivity as a risk of low reliability in our case. However, the pertinence of expert opinion in validity achievement is exceedingly acknowledged [20]. In addition, there is a potential risk of bias in the selection of experts; we have attempted to overcome this by choosing experts of different fields (that is, gynaecology, hysteroscopy, medical education, and virtual-reality simulation) affiliated with three different hospital departments and two different simulation centres.

Another validity source was *internal structure*; high reliability above or equal to 0.80 was found for all of the metrics in the six modules except for resected pathology. For high-stakes summative tests the preferred reliability should be between 0.80 and 0.89 in order to guarantee a satisfactory consistency and reproducibility of the test scores [21]. Hence, the interpretation of the score achieved in the metric resected pathology with a Cronbach's alpha of 0.58 is not as reliable as the other metrics. This metric should therefore not solely serve as assessment of performance. Nevertheless, considering resection as imperative in the clinic, the metric was retained in the overall score.

Performances between the three experience groups were significantly different both with regard to performance on the specific metrics and the Overall Scores (validity evidence regarding *relationship to other variables*, that is, test score). A somewhat unpredicted outcome was found when comparing the group performances with each other; the residents did neither significantly vary from the medical students nor from the experienced gynaecologists. We consider this limitation as a type II error and can therefore not draw genuine inferences hereupon. In reality, we believe that substantial differences exist between the three groups' skills that could have been uncovered with sufficient statistical power, that is, larger sample size. Another proposed explanation is that experienced physicians—with many years of practice in the operating theatre—underperform on a simulator due to its absence of haptic and visual feedback [22]. We cannot rule out that this contributes to our findings. Nonetheless, hysteroscopy experience was shown to correlate significantly with the participants' overall scores indicating that long-term practice pays off on performance. The superior performance of experts regardless of domain is acknowledged as consistent due to several years of deliberate practice [23].

An imperative asset of a test used in high-stakes exams and for certification is a credible pass/fail score due to its pivotal *consequences* [17]. It is expected that one can rely on the test content to fairly discriminate between the proficient participants and those that are not. However, no gold standard for determining such score exists and no ideal score can guarantee a clear-cut separation of the qualified from the non-qualified [17]. The Contrasting Groups' method has been applied in several other studies concerning technical skills acquisition on simulators [24]. The pass/fail score of our test largely illustrates the contrast between the performances of the medical students and the experienced gynaecologists (Fig. 3).

Three of the experienced gynaecologists failed the test, though their experiences in hysteroscopy did not vary from the other gynaecologists. As already mentioned, a degree of unfamiliarity with virtual-reality simulation could explain their performances. Lack of realism has in some studies been reported as a reason for lower expert ratings of virtual-reality simulators [25, 26]. Although an assorted panel of experts deemed the test content as relevant in our study, a record of the participants' views on the simulator-based test could have clarified issues on simulator fidelity as a potential reason for underperformance. We recognize this as a limitation.

The population at aim for our test is the residents of whom 9 out of 17 passed. Shifting the pass/fail score to a higher score would inevitably make the test more challenging for the residents offering them the benefit of additional training in their initial stages of technical skills acquisition [27]. Within the realm of mastery learning, this pass/fail score will function as a reasonable benchmark for the residents to *master* the basic skills of hysteroscopy [13, 27].

Furthermore, is it fundamental to emphasise that our test exclusively quantifies technical aspects of the hysteroscopic procedure via simulator metrics. It must be kept in mind that other aspects are at least of equal importance in assessing competence [28]. Some of these aspects are the pre- and postoperative clinical knowledge base and reasoning (for instance indications and complications) as well as the non-technical skills associated with the procedure [29]. These aspects must be evaluated separately.

With this study, we have developed a standardised, simulation-based test of competence in hysteroscopy and established arguments of validity evidence for the test. The test could be used to set up a simulation-based mastery learning training program and as an aid to decide when trainees have acquired basic competence in hysteroscopy.

Acknowledgements The authors thank senior consultant, Ivan Moulun Grunnet, an experienced hysteroscopist, for his contribution to the test contents.

Funding The study did not receive any funding or financial support.

Compliance with ethical standards

Disclosures The authors Mona M. Savran, Anders Bo Nielsen, Bente Bækholm Poulsen, Poul Bak Thorsen, and Lars Konge have no conflicts of interest or financial ties to disclose.

Ethical approval Ethical approval was granted by the local ethics committee (S-20172000-10) on 16 January 2017.

Informed consent Written informed consent was obtained from all participants before enrolment.

References

1. Mencaglia L, de Albuquerque Neto LC, Alvarez RAA (2013) Manual of hysteroscopy—diagnostic, operative and office hysteroscopy. Manual of hysteroscopy Endo Press, Tuttlingen
2. Clark TJ, Voit D, Gupta JK, Hyde C, Song F, Khan KS (2002) Accuracy of hysteroscopy in the diagnosis of endometrial cancer and hyperplasia: a systematic quantitative review. JAMA 288:1610–1621
3. Elessawy M, Skrzipczyk M, Eckmann-Scholz C, Maass N, Mettler L, Guenther V, van Mackelenbergh M, Bauerschlag DO, Alkatout I (2017) Integration and validation of hysteroscopy simulation in the surgical training curriculum. J Surg Educ 74:84–90
4. Korndorffer JR, Hayes DJ, Dunne JB, Sierra R, Touchard CL, Markert RJ, Scott DJ (2005) Development and transferability of a cost-effective laparoscopic camera navigation simulator. Surg Endosc 19:161–167
5. McGurgan PM, McIlwaine P (2015) Complications of hysteroscopy and how to avoid them. Best Pract Res Clin Obstet Gynaecol 29:982–993
6. Rackow BW, Solnik MJ, Tu FF, Senapati S, Pozolo KE, Du H (2012) Deliberate practice improves obstetrics and gynecology residents' hysteroscopy skills. J Grad Med Educ 4:329–334
7. Savran MM, Sørensen SMD, Konge L, Tolsgaard MG, Bjerrum F (2016) Training and assessment of hysteroscopic skills: a systematic review. J Surg Educ 73:906–918
8. Hiemstra E, Kolkman W, Jansen FW (2008) Skills training in minimally invasive surgery in Dutch obstetrics and gynecology residency curriculum. Gynecol Surg 5:321–325
9. De Wilde RL, Hucke J, Kolmorgen K, Tinneberg H (2011) Recommendations by the Gynecologic Endoscopy Working Group of the German Society of Obstetrics and Gynecology for the advancement of training and education in minimal-access surgery. Arch Gynecol Obstet 283:509–512
10. Cook DA, Hatala R, Brydges R, Zendejas B, Szostek JH, Wang AT, Erwin PJ, Hamstra SJ (2011) Technology-enhanced simulation for health professions education: a systematic review and meta-analysis. JAMA 306:978–988
11. Cook DA, Brydges R, Zendejas B, Hamstra SJ, Hatala R (2013) Mastery learning for health professionals using technology-enhanced simulation: a systematic review and meta-analysis. Acad Med 88:1178–1186
12. Cohen ER, McGaghie WC, Wayne DB, Lineberry M, Yudkowsky R, Barsuk JH (2015) Recommendations for reporting mastery education research in medicine (ReMERM). Acad Med 90:1509–1514
13. Lineberry M, Soo Park Y, Cook DA, Yudkowsky R (2015) Making the case for mastery learning assessments: key issues in validation and justification. Acad Med 90:1445–1450
14. Bloch R, Norman G (2012) Generalizability theory for the perplexed: a practical introduction and guide: AMEE Guide No. 68. Med Teach 34:960–992
15. Downing SM, Haladyna TM (2009) Validity and its threats. In: Yudkowsky R, Downing SM (eds) Assessment in health professions education. Routledge, New York, pp 21–55
16. Borgersen NJ, Naur TMH, Sørensen SMD, Bjerrum F, Konge L, Subhi Y, Thomsen ASS (2018) Gathering validity evidence for surgical simulation: a systematic review. Ann Surg 267:1063–1068
17. Yudkowsky R, Downing SM, Tekian A (2009) Standard setting. In: Yudkowsky R, Downing SM (eds) Assessment in health professions education. Routledge, New York, pp 119–146
18. Jensen JT, Savran MM, Møller AM, Vilmann P, Hornslet P, Konge L (2016) Development and validation of a theoretical test in non-anaesthesiologist-administered propofol sedation for gastrointestinal endoscopy. Scand J Gastroenterol 51:872–879
19. Hasson F, Keeney S, McKenna H (2000) Research guidelines for the Delphi survey technique. J Adv Nurs 32:1008–1015
20. Streiner DL, Norman GR (2008) Divising the items. In: Streiner DL, Norman GR (eds) Health measurement scales: a practical guide to their development and use. Oxford University Press, Oxford, pp 18–38
21. Downing SM (2004) Reliability: on the reproducibility of assessment data. Med Educ 38:1006–1012
22. Moore AK, Grow DR, Bush RW, Seymour NE (2008) Novices outperform experienced laparoscopists on virtual reality laparoscopy simulator. JSLS 12:358–362
23. Ericsson KA (2004) Deliberate practice and the acquisition and maintenance of expert performance in medicine and related domains. Acad Med 79:S70–S81
24. Madsen ME, Konge L, Nørgaard LN, Tabor A, Ringsted C, Klemmensen AK, Ottesen B, Tolsgaard MG (2014) Assessment of performance measures and learning curves for use of a virtual-reality ultrasound simulator in transvaginal ultrasound examination. Ultrasound Obstet Gynecol 44:693–699
25. Alsalamah A, Campo R, Tanos V, Grimbizis G, Van Belle Y, Hood K, Pugh N, Amso N (2017) Face and content validity of the virtual reality simulator 'ScanTrainer®'. Gynecol Surg 14:18

26. Chalasani V, Cool DW, Sherebrin S, Fenster A, Chin J, Izawa JI (2011) Development and validation of a virtual reality transrectal ultrasound guided prostatic biopsy simulator. Can Urol Assoc J 5:19–26
27. Yudkowsky R, Park YS, Lineberry M, Knox A, Ritter EM (2015) Setting mastery learning standards. Acad Med 90:1495–1500
28. Miller GE (1990) The assessment of clinical skills/competence/performance. Acad Med 65:S63–S67
29. Gjeraa K, Spanager L, Konge L, Petersen RH, Østergaard D (2016) Non-technical skills in minimally invasive surgery teams: a systematic review. Surg Endosc 30:5185–5199